16 July 2004

PERFECT SIMULATION OF HAWKES PROCESSES

JESPER MØLLER,* *Aalborg University*

JAKOB G. RASMUSSEN,** *Aalborg University*

Abstract

This article concerns a perfect simulation algorithm for unmarked and marked Hawkes processes. The usual straightforward simulation algorithm suffers from edge effects, whereas our perfect simulation algorithm does not. By viewing Hawkes processes as Poisson cluster processes and using their branching and conditional independence structure, useful approximations of the distribution function for the length of a cluster are derived. This is used to construct upper and lower processes for the perfect simulation algorithm. Examples of applications and empirical results are presented.

Keywords: Approximate simulation; dominated coupling from the past; edge effects; exact simulation; Hawkes process; marked Hawkes process; marked point processes; perfect simulation; point processes; Poisson cluster processes; thinning; upper and lower processes

AMS 2000 Subject Classification: Primary 60G55
Secondary 68U20

1. Introduction

This paper concerns a perfect (or exact) simulation algorithm for unmarked and marked Hawkes processes [7, 9, 10, 11, 13]. Such processes play a fundamental role for point process theory and its applications, cf., for example, p. 183 in [7]. Particularly, marked Hawkes processes have applications in seismology [12, 21, 22, 26] and neurophysiology [3, 6].

* Postal address: Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7G, DK-9220 Aalborg, Denmark. Email address: jm@math.auc.dk

** Postal address: As above. Email address: jgr@math.auc.dk

For mathematical convenience (see Sections 2.2 and 6) we restrict attention to marked Hawkes processes with unpredictable marks. Such a process $X = \{(t_i, Z_i)\} \subset \mathbb{R} \times M$ is defined as follows, where M denotes an arbitrary mark space (equipped with a σ-field).

Definition of a marked Hawkes process with unpredictable marks:

- Each event (or time) t_i is of one of two types: an *immigrant* or an *offspring*. The immigrants follow a Poisson point process, with an intensity function $\mu(t)$ on $\mathbb{R}$, which is locally integrable. This process is called the *immigrant process*.
- The process has *unpredictable marks* in the sense that each mark Z_i follows the same probability distribution Q on M, which is independent of t_i and the previous history $\{(t_k, Z_k) : t_k < t_i\}$.
- If we condition on (t_i, Z_i), then independently of the previous history, (t_i, Z_i) generates a Poisson point process $\Phi(t_i)$ of offspring on (t_i, ∞), with intensity function $\gamma_i(t) = \gamma(t - t_i, Z_i)$, $t > t_i$, where γ is a non-negative measurable function on $(0, \infty) \times M$ (for convenience we suppress in the notation that $\Phi(t_i)$ may depend on Z_i). The process $\Phi(t_i)$ is called an *offspring process*, and we refer to γ_i and γ as *fertility rates*.

The paper is organised as follows. Section 2 contains some preliminaries, in particular we view the marked Hawkes process as a Poisson cluster process and describe the branching and conditional independence structure of each cluster. Section 3 describes a straightforward simulation algorithm, which suffers from edge effects, and a perfect simulation algorithm, which does not. The perfect simulation algorithm is derived using similar principles as in Brix and Kendall [5], but our algorithm requires the knowledge of the cumulative distribution function (c.d.f.) F for the length of a cluster. Section 4 determines an integral equation for F by generalising a result in Hawkes and Oakes [13]. Particularly, it is discussed how to approximate F (since a closed form expression is unknown) by establishing certain monotonicity and convergence results. Section 5 completes the perfect simulation algorithm by using a dominating process and upper and lower processes in a similar fashion as in the Propp-Wilson algorithm [23], or rather as in the dominated coupling from the past algorithm by Kendall and Møller [15]. Moreover, throughout Sections 2–5, illuminating examples

and empirical results are presented. Finally, Section 6 contains a discussion of our algorithms and results and how to extend these to more general settings.

2. Preliminaries and examples

2.1. Some useful properties

From the definition above the following properties immediately follow. The marks are i.i.d. with distribution Q. In the special case where $\gamma(t, z) = \gamma(t)$ does not depend on its second argument, the events follow an unmarked Hawkes process. Apart from that case, the events and the marks are dependent processes. The conditional intensity function $\lambda(t)$ at time $t \in \mathbb{R}$ for the events given the previous history $\{(t_k, Z_k) : t_k < t\}$ (see e.g. [7]) is given by

$$\lambda(t) = \mu(t) + \sum_{t_i < t} \gamma(t - t_i, Z_i). \tag{1}$$

2.2. The branching and conditional independence structure of marked Hawkes processes

It becomes useful to view the marked Hawkes process as a Poisson cluster process, with cluster centres given by the immigrants, and clusters defined as follows. For events $t_i < t_j$, we say that (t_j, Z_j) has *ancestor* t_i *of order* $n \geq 1$ if there is a sequence $s_1 \ldots, s_n$ of offspring such that $s_n = t_j$ and $s_k \in \Phi(s_{k-1})$ for $k = 1, \ldots, n$, where $s_0 = t_i$. We say then that t_j is an *offspring of* n*-th generation with respect to* t_i; for convenience we say that t_i is of zeroth generation with respect to itself. Now, define the *total offspring process* C_i as all (t_j, Z_j) such that t_j is an event of generation $n \in \mathbb{N}_0$ with respect to t_i (note that $(t_i, Z_i) \in C_i$). The *clusters* are defined as those C_i where t_i is an immigrant.

The *total offspring processes have the same probability structure relative to their generating events* because of the following *branching structure* (see also Figure 1).

- Conditional on an event t_i, its mark Z_i follows Q (independently of the previous history); if we also condition on Z_i, then $\Phi(t_i)$ (the first generation of offspring with respect to t_i) is a Poisson process with intensity function $\gamma_i(t)$; conditioning further on $\Phi(t_i)$, the events in $\Phi(t_i)$ generate independent total offspring processes C_j, $t_j \in \Phi(t_i)$.

- If we furthermore condition on such a t_j, then Z_j follows Q; if we also condition on Z_j, then $\Phi(t_j)$ (the second generation of offspring with respect to t_i) is a Poisson process with intensity function $\gamma_j(t)$; conditioning moreover on $\Phi(t_j)$, the events in $\Phi(t_j)$ generate independent total offspring processes C_k, $t_k \in \Phi(t_k)$.
- Similarly for the third, fourth, ... generation of offspring with respect to t_i.

Since $\gamma_i(t) = \gamma(t - t_i, Z_i)$ for any event t_i, we see that conditional on events $t_i < t_j$, the translated total offspring processes $C_i - t_i \equiv \{(t_l - t_i, Z_l) : (t_l, Z_l) \in C_i\}$ and $C_j - t_j \equiv \{(t_l - t_j, Z_l) : (t_l, Z_l) \in C_j\}$ are identically distributed.

In particular, conditional on the immigrants, the clusters relative to their cluster centres (the immigrants) are i.i.d. with distribution P, say. Furthermore, conditional on the n-th generation events $\mathcal{G}_n$, say, in a cluster, the translated total offspring processes $C_j - t_j$ with $t_j \in \mathcal{G}_n$ are i.i.d. with distribution P. We refer to this last property as the i.i.d. self-similarity property of offspring processes or for short the *self-similarity property*. Note that the assumption of unpredictable marks is essential for these properties to hold.

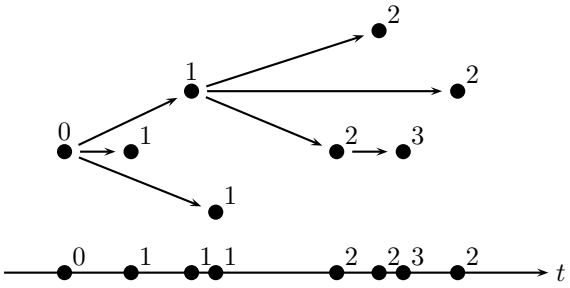

FIGURE 1: Top: The branching structure of a cluster in an unmarked case; the numbers indicate which generations the events are. Bottom: The events on the time axis.

2.3. A basic assumption and some terminology and notation

Let F denote the c.d.f. for the length L of a cluster, i.e. the time between the immigrant and the last event of the cluster. Consider the mean number of events in

any offspring process $\Phi(t_i)$, $\bar{\nu} \equiv \mathrm{E}\nu$, where

$$\nu = \int_0^\infty \gamma(t, Z)\,\mathrm{d}t$$

is the *total fertility rate of an offspring process* and Z denotes a *generic mark* with distribution Q. Henceforth we assume that

$$0 < \bar{\nu} < 1. \tag{2}$$

The condition $\bar{\nu} < 1$ appears commonly in the literature on Hawkes processes, see e.g. [4, 7, 13]. It implies that

$$F(0) = \mathrm{E}\mathrm{e}^{-\nu} > 0 \tag{3}$$

where $F(0)$ is the probability that a cluster has no offspring. It is equivalent to assuming that $\mathrm{E}S < \infty$, where S denotes the number of events in a cluster: By induction on $n \in \mathbb{N}_0$, because of the branching and conditional independence structure of a cluster, $\bar{\nu}^n$ is the mean number of generation n events in a cluster, so

$$\mathrm{E}S = 1 + \bar{\nu} + \bar{\nu}^2 + \cdots = 1/(1 - \bar{\nu}) \tag{4}$$

if $\bar{\nu} < 1$, while $\mathrm{E}S = \infty$ otherwise.

The other condition $\bar{\nu} > 0$ excludes the trivial case where there are almost surely no offspring. It is readily seen to be equivalent to

$$F < 1. \tag{5}$$

Furthermore,

$$h(t) = \mathrm{E}[\gamma(t, Z)/\nu], \qquad t > 0, \tag{6}$$

and

$$\bar{h}(t) = \mathrm{E}\gamma(t, Z)/\bar{\nu}, \qquad t > 0, \tag{7}$$

are well-defined densities (with respect to Lebesgue measure). The density $\bar{h}$ will play a keyrole later in this paper; it can be interpreted as the *normalised intensity function for the first generation of offspring in a cluster started at time* 0. Note that h specifies the *density of the distance R from an arbitrary offspring to its nearest ancestor*. In the sequel, since the clusters relative to their cluster centers are i.i.d. (Section 2.2), we

assume without loss of generality that L, R and S are defined with respect to the same immigrant $t_0 = 0$, with mark $Z_0 = Z$.

Clearly, if $L > 0$ then $R > t$ implies $L > t$, so the distribution of L has a thicker tail than that of R. The probability function for S is given by $P(S = k) = P(S_{n+1} = k-1|S_n = k)/k$, $k \in \mathbb{N}$, where S_n denotes the number of events of n-th generation and $n \in \mathbb{N}$ is arbitrary (see [8] or Theorem 2.11.2 in [14]). Thus

$$\mathrm{P}(S = k) = \mathrm{E}\left[\mathrm{e}^{-k\nu}(k\nu)^{k-1}/k!\right], \qquad k \in \mathbb{N}. \tag{8}$$

2.4. Examples

Throughout the paper, we illustrate the results with the following cases.

Example 1. (*Unmarked process*) An unmarked Hawkes process with exponentially decaying fertility rate is given by

$$\bar{\nu} = \nu = \alpha, \qquad \bar{h}(t) = h(t) = \beta \mathrm{e}^{-\beta t},$$

where $0 < \alpha < 1$ and $\beta > 0$ are parameters. Here $1/\beta$ is a scale parameter for both the distribution of R and the distribution of L.

Figure 2 (at left) shows perfect simulations on $[0, 10]$ of this process when $\mu(t) = 1$ is constant, $\alpha = 0.9$, and $\beta = 10, 5, 2, 1$. By (4), we expect to see about 10 clusters (in total) and 100 events. The clusters of course become more visible as β increases.

Figure 3 (at left) shows six simulations of clusters with $\alpha = 0.9$ (being an inverse scaling parameter, β is irrelevant since we have omitted showing the scale to get comparable results for this example and the following two examples). All the clusters have been simulated conditional on $S > 1$ to avoid the frequent and rather uninteresting case containing only the immigrant. These few simulations indicate the general tendency that L vary fairly much. □

Example 2. (*Birth-death process*) Consider a marked Hawkes process with

$$\gamma(t, Z) = \alpha \mathbf{1}[t \le Z]/\mathrm{E}Z,$$

where $0 < \alpha < 1$ is a parameter, Z is a positive random variable, and $\mathbf{1}[\cdot]$ denotes the indicator function. Then X can be viewed as a birth and death process, with birth at

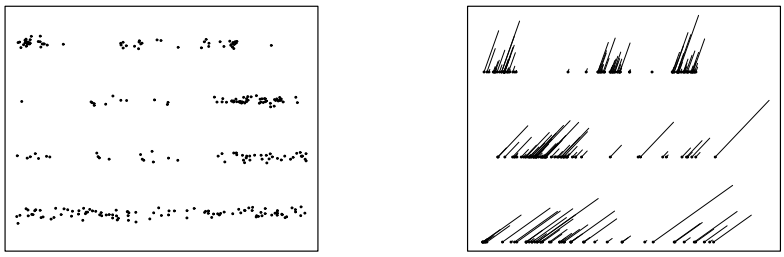

FIGURE 2: Left: Four perfect simulations on $[0, 10]$ of the unmarked Hawkes process (Example 1) with parameters $\alpha = 0.9$, $\mu = 1$, and $\beta = 10, 5, 2, 1$ (top to bottom). Random jitter has been added in the vertical direction to help distinguishing events located close together. Right: Three perfect simulations on $[0, 10]$ of the birth-death Hawkes process (Example 2) with parameters $\alpha = 0.9$, $\mu = 1$, and $\beta = 5, 2, 1$ (top to bottom), where the projections of the lines onto the horizontal axis show the size of the marks.

time t_i and survival time Z_i of the i'th individual. The birth rate is

$$\lambda(t) = \mu(t) + (\alpha/\mathrm{E}Z)\mathrm{card}\big(\{i : t_i < t \leq t_i + Z_i\}\big), \qquad t \in \mathbb{R},$$

cf. (1). Moreover,

$$\nu = \alpha Z/\mathrm{E}Z, \qquad \bar{\nu} = \alpha, \qquad h(t) = \mathrm{E}(\mathbf{1}[t \leq Z]/Z), \qquad \bar{h}(t) = \mathrm{P}(Z \geq t)/\mathrm{E}Z.$$

Since ν is random, the distribution of S is more dispersed than in the unmarked case, cf. (8).

The special case where $\mu(t) = \mu$ is constant and Z is exponentially distributed with mean $1/\beta$ is considered at page 136 in [4]. Then X is a time-homogeneous Markov birth-death process with birth rate $\mu + \alpha\beta n$ and death rate βn, when n is the number of living individuals. In this case $\bar{h}(t) = \beta \mathrm{e}^{-\beta t}$ and $h(t) = \beta E_1(\beta t)$, where $E_1(s) = \int_s^\infty \mathrm{e}^{-t}/t \, \mathrm{d}t$ is the exponential integral function. As in Example 1, $1/\beta$ is a scale parameter for the distribution of L. As discussed in Example 8 in Section 5, the stationary distribution (i.e. the distribution of X at any fixed time) is known up to proportionality and it is possible to simulate from this by rejection sampling.

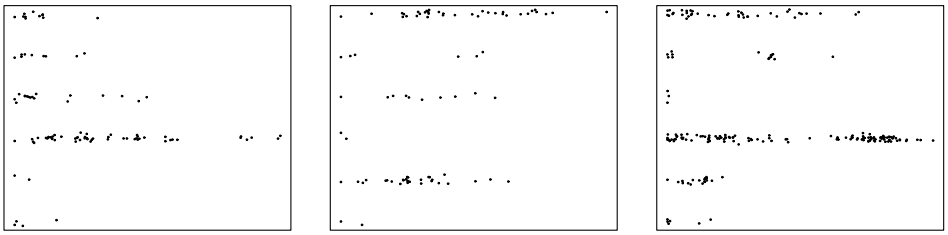

FIGURE 3: Left: Six simulations of clusters started at zero and conditioned on $S > 1$ in the unmarked case with $\alpha = 0.9$. Middle: As left, but for the birth-death case. Right: As left, but for the heavy-tailed case. Different scalings are used in the three cases.

Figure 2 (at right) shows three perfect simulations in the Markov case on $[0, 10]$ with $\mu = 1$, $\alpha = 0.9$, and $\beta = 5, 2, 1$, where the marks are indicated by line segments of different lengths. Figure 3 (at middle) shows six simulations of clusters (with marks excluded) with $\alpha = 0.9$ simulated conditional on $S > 1$. These simulations slightly indicate that L is more dispersed than in Example 1, since the marks introduce additional variation in the cluster lengths. In fact the coefficient of variation estimated from 10000 perfect simulations is 1.92 for Example 1 and 2.85 for the present case. □

Example 3. (*A heavy-tailed distribution for* L) Suppose that

$$\gamma(t, Z) = \alpha Z \mathrm{e}^{-tZ},$$

where $\alpha \in (0, 1)$ is a parameter and Z is exponentially distributed with mean $1/\beta$. Then $\bar{\nu} = \nu = \alpha$ is constant, so the distribution of S is the same as in the unmarked case, cf. (8). Further,

$$h(t) = \bar{h}(t) = \beta/(t + \beta)^2$$

specifies a Pareto density. This is a heavy-tailed distribution as is has infinite Laplace transform ($\mathcal{L}(\theta) = \mathrm{E}\mathrm{e}^{\theta R} = \infty$ for all $\theta > 0$). Moreover it has infinite moments ($\mathrm{E}\big(R^p\big) = \infty$ for all $p \geq 1$). Consequently, L also has a heavy-tailed distribution with infinite moments and infinite Laplace transform. Note that β is a scale parameter for the distribution of L.

Figure 3 (at right) shows six simulations of clusters with $\alpha = 0.9$ and $\beta = 1$. These indicate that L is much more dispersed than in Examples 1 and 2 (in fact the dispersion is infinite in the present case). □

3. Perfect Simulation

In this section we introduce the usual straightforward simulation algorithm, which suffers from edge effects, and our perfect simulation algorithm, which solves this problem.

3.1. Approximate simulation and edge effects

The general approach for simulating a (marked or unmarked) point process is to use a thinning algorithm such as Shedler-Lewis thinning algorithm or Ogata's modified thinning algorithm, see e.g. [7]. However, for a (marked or unmarked) Hawkes process, the easiest approach is to generate it as a Poisson cluster process as in the following algorithm.

Algorithm 1. The following steps (i)-(iii) generate a simulation of those $(t_i, Z_i) \in X$ with $0 \le t_i < t_+$, where $t_+ \in (0, \infty]$ is a user-specified parameter.

(i) Simulate the immigrants on $[t_-, t_+)$ where $t_- \in [-\infty, 0]$ is a user-specified parameter.

(ii) For each such immigrant t_i, simulate Z_i and those $(t_j, Z_j) \in C_i$ with $t_i < t_j < t_+$.

(iii) The output is all marked points (t_j, Z_j) with $t_j \in [0, t_+)$.

In principle steps (i) and (ii) are easy because the immigrants follow a Poisson process, and because of the branching construction of clusters into Poisson processes, cf. Section 2.2. However, ideally we should take $t_- = -\infty$, but in practice we need to determine t_- such that $\int_{t_-}^{0} \mu(t)\,\mathrm{d}t < \infty$. When $\int_{-\infty}^{t_-} \mu(t)\,\mathrm{d}t > 0$, Algorithm 1 suffers from edge effects, since clusters generated by immigrants before time t_- may contain offspring in $[0, t_+)$. In [17] this algorithm is investigated much more thoroughly and various measures for the edge effects are introduced; in the rest of this article, however, we will consider how to simulate the marked Hawkes process without edge effects.

3.2. Description of perfect simulation algorithm

Assuming for the moment that F (the c.d.f. for the length of a cluster) is known, the following algorithm for perfect simulation of the marked Hawkes process is similar to the simulation of Poisson cluster processes without edge effects given in Brix and Kendall [5]; see also [16, 20].

Algorithm 2. Let I_1 be the point process of immigrants on $[0, t_+)$, and let I_2 be the point process of immigrants $t_i < 0$ such that $\{(t_j, Z_j) \in C_i : t_j \in [0, \infty)\} \neq \emptyset$.

(i) Simulate I_1 as a Poisson process with intensity function $\lambda_1(t) = \mu(t)$ on $[0, t_+)$.

(ii) For each $t_i \in I_1$, simulate Z_i and those $(t_j, Z_j) \in C_i$ with $t_i < t_j < t_+$.

(iii) Simulate I_2 as a Poisson process with intensity function $\lambda_2(t) = (1 - F(-t))\mu(t)$ on $(-\infty, 0)$.

(iv) For each $t_i \in I_2$, simulate Z_i and $\{(t_j, Z_j) \in C_i : t_j \in [0, t_+)\}$ conditional on that $\{(t_j, Z_j) \in C_i : t_j \in [0, \infty)\} \neq \emptyset$.

(v) The output is all marked points from (i), (ii), and (iv).

Remark 1. In steps (i) and (ii) of Algorithm 2, we use Algorithm 1 (with $t_- = 0$). In step (iv), it is not obvious how to construct an elegant approach ensuring that at least one point will fall after 0. Instead we use a simple rejection sampler: we repeatedly simulate Z_i from Q and the successive generations of offspring t_j to t_i (together with their marks Z_j) until there is at least one event of C_i after time 0.

The key point is how to simulate I_2 in step (iii), since this requires the knowledge of F, which is unknown on closed from (Remark 3, Section 4.1). In Section 4 we address this problem, and in Section 5 we construct an algorithm for simulating I_2. In practice we must require that I_2 is (almost surely) finite or equivalently that

$$\int_{-\infty}^{0} (1 - F(-t))\mu(t)\,\mathrm{d}t < \infty. \tag{9}$$

In the case where $\mu(t)$ is bounded, (9) is satisfied if $\sup_{t \geq 0} \mu(t)\,\mathrm{E}L < \infty$. A condition for finiteness of $\mathrm{E}L$ is established in Lemma 1 and Remark 2 below. □

Proposition 1. *The output of Algorithm 2 follows the distribution of the marked Hawkes process.*

Proof. The immigrant process minus $I_1 \cup I_2$ generates clusters with no events in $[0, t_+)$. Since I_1 is the immigrants on $[0, t_+)$, it follows directly that I_1 is a Poisson process with intensity $\lambda_1(t) = \mu(t)$ on $[0, t_+)$. Since I_2 is those immigrants on $(-\infty, 0)$ with offspring after 0, I_2 can be viewed as an independent thinning of the immigrant process with retention probability $p(t) = 1 - F(-t)$, and thus I_2 is a Poisson process with intensity $\lambda_2(t) = (1 - F(-t))\mu(t)$. Since I_1 and I_2 are independent, it follows from Section 2.2 that $\{C_i : t_i \in I_1\}$ and $\{C_i : t_i \in I_2\}$ are independent. Viewing the marked Hawkes process as a Poisson cluster process, it follows from Remark 1 that the clusters are generated in the right way in (ii) and (iv) when we only want to sample those marked points (t_j, Z_j) with $t_j \in [0, t_+)$. Thus Algorithm 2 is correct. □

Using a notation as in Section 2.3, the following lemma generalises and sharpens a result in [13] about the mean length of a cluster.

Lemma 1. *We have that*

$$\frac{1}{\mathrm{E}\mathrm{e}^{-\nu}}\mathrm{E}\left[(1 - \mathrm{e}^{-\nu})\mathrm{E}[R|Z]\right] \leq \mathrm{E}L \leq \frac{\bar{\nu}}{1 - \bar{\nu}}\mathrm{E}\bar{R}. \tag{10}$$

Proof. Consider a cluster starting with an immigrant at time $t_0 = 0$, with mark $Z_0 = Z$, cf. Section 2.2. For $t_j \in \mathcal{G}_1$, let R_j denote the distance from t_j to 0, and L_j the length of the total offspring C_j process started by t_j. Then $L = \max\{R_j + L_j : t_j \in \mathcal{G}_1\}$, so if we condition on Z, and let $R_{j,z}$ be distributed as R_j conditional on $Z = z$, then

$$\mathrm{E}L = \mathrm{E}\mathrm{E}[L|Z] = \mathrm{E}\left[\sum_{i=1}^{\infty} \frac{\mathrm{e}^{-\nu}\nu^i}{i!}\mathrm{E}\left[\max\{R_{j,Z} + L_j : j = 1, \ldots, i\}\right]\right]. \tag{11}$$

To obtain the upper inequality observe that

$$\mathrm{E}L \leq \mathrm{E}\left[\sum_{i=1}^{\infty} \frac{\mathrm{e}^{-\nu}\nu^i}{i!}\mathrm{E}\left[\sum_{j=1}^{i}(R_{j,Z} + L_j)\right]\right] = \mathrm{E}[\nu\mathrm{E}[R|Z]] + \bar{\nu}\mathrm{E}L,$$

where we have used that the L_j are identically distributed and has the same distribution as L because of the self-similarity property (Section 2.2), and that the R_j are identically distributed when conditioned on Z. Hence

$$\mathrm{E}L \leq \frac{1}{1 - \bar{\nu}}\mathrm{E}[\nu\mathrm{E}[R|Z]] = \frac{1}{1 - \bar{\nu}}\mathrm{E}\left[\int_0^{\infty} s\gamma(s, Z)\mathrm{d}s\right] = \frac{\bar{\nu}}{1 - \bar{\nu}}\mathrm{E}\bar{R},$$

which verifies the upper inequality. Finally, by (11),

$$\mathrm{E}L \geq \mathrm{E}\left[\sum_{i=1}^{\infty} \frac{\mathrm{e}^{-\nu}\nu^i}{i!}(\mathrm{E}[R|Z] + \mathrm{E}L)\right] = \mathrm{E}\left[(1 - \mathrm{e}^{-\nu})\mathrm{E}[R|Z]\right] + \mathrm{E}[1 - \mathrm{e}^{-\nu}]\mathrm{E}L,$$

which reduces to the lower inequality. □

Remark 2. If either ν or γ/ν is independent of Z (in other words, either the number or the locations of offspring in an offspring process are independent of the mark associated to the generic event), then it is easily proven that $\bar{h} = h$ and thus (10) reduces to

$$\left(\frac{1}{\mathrm{E}\mathrm{e}^{-\nu}} - 1\right) \mathrm{E}R \le \mathrm{E}L \le \frac{\bar{\nu}}{1-\bar{\nu}} \mathrm{E}R.$$

Consequently, $\mathrm{E}L < \infty$ if and only if $\mathrm{E}R < \infty$. This immediately shows that $\mathrm{E}L < \infty$ in Example 1 and $\mathrm{E}L = \infty$ in Example 3. In Example 2 when Z is exponentially distributed with mean $1/\beta$, (10) becomes $\alpha(\alpha+2)/(2(\alpha+1)\beta) \le \mathrm{E}L \le \alpha/(\beta(1-\alpha))$, so in this case $\mathrm{E}L < \infty$. Not surprisingly, apart from small values of $\alpha \in (0,1)$, the bounds are rather poor and of little use except in establishing finiteness of $\mathrm{E}L$. □

4. The distribution of the length of a cluster

In this section we derive various distributional results concerning the length L of a cluster. The results are needed in Section 5 to complete step (iii) in Algorithm 2; however, many of the results are also of own interest.

4.1. An integral equation for F

Below in Proposition 2 an integral equation for F is derived, and it is discussed how to approximate F by numerical methods, using a certain recursion. Proposition 2 is a generalisation of Theorem 5 in Hawkes and Oakes [13], which is proved using void probabilities obtained from a general result for the probability generating functional for an unmarked Hawkes process. However, as Daley and Vere-Jones [7] point out, the probability generating functional for the marked Hawkes process is difficult to obtain. We give a direct proof based on void probabilities.

For $n \in \mathbb{N}_0$, let 1_n denote the c.d.f. for the length of a cluster when all events of generation $n+1, n+2, \ldots$ are removed (it becomes clear in Section 4.2 why we use the notation 1_n). Clearly, 1_n is decreasing in n, $1_n \to F$ pointwise as $n \to \infty$, and

$$1_0(t) = 1, \qquad t \ge 0. \tag{12}$$

Furthermore, let $\mathcal{C}$ denote the class of Borel functions $f : [0,\infty) \mapsto [0,1]$. For $f \in \mathcal{C}$,

define $\varphi(f) \in \mathcal{C}$ by

$$\varphi(f)(t) = \mathrm{E}\left[\exp\left(-\nu + \int_0^t f(t-s)\gamma(s,Z)\,\mathrm{d}s\right)\right], \qquad t \geq 0. \tag{13}$$

Proposition 2. *We have that*

$$1_n = \varphi(1_{n-1}), \qquad n \in \mathbb{N}, \tag{14}$$

and

$$F = \varphi(F). \tag{15}$$

Proof. As in the proof of Lemma 1, we can consider a cluster started at time $t_0 = 0$ with associated marks $Z_0 = Z$. For fixed $t \geq 0$ and $n \in \mathbb{N}$, split $\Phi(0)$ into three point processes Φ_1, Φ_2, Φ_3: Φ_1 consists of those first generation offspring $t_i \in \Phi(0) \cap [0,t)$ which do not generate events of generation $n-1$ or lower with respect to t_i on $[t,\infty)$; $\Phi_2 = (\Phi(0) \cap [0,t)) \setminus \Phi_1$ consists of the remaining first generation offspring on $[0,t)$; and $\Phi_3 = \Phi(0) \cap [t,\infty)$ are the first generation offspring on $[t,\infty)$. Conditional on Z, we have that Φ_1, Φ_2, and Φ_3 are independent Poisson processes with intensity functions $\lambda_1(s) = \gamma(s,Z)F_{n-1}(t-s)$ on $[0,t)$, $\lambda_2(s) = \gamma(s,Z)(1-F_{n-1}(t-s))$ on $[0,t)$, and $\lambda_3(s) = \gamma(s,Z)$ on $[t,\infty)$, respectively. This follows by an independent thinning argument, since conditional on $\mathcal{G}_n$ (the n-th generation of offspring in C_0), the processes $C_j - t_j$ with $t_j \in \mathcal{G}_n$ are i.i.d. and distributed as C_0 (this is the self-similarity property from Section 2.2). Consequently,

$$1_n(t) = \mathrm{E}[\mathrm{P}(\Phi_2 = \emptyset|Z)\mathrm{P}(\Phi_3 = \emptyset|Z)] = \mathrm{E}\exp\left(-\int_0^t \lambda_2(s,Z)\,\mathrm{d}s - \int_t^\infty \lambda_3(s,Z)\,\mathrm{d}s\right)$$

which reduces to (14). Taking the limit as $n \to \infty$ on both sides of (14), we obtain (15) by monotone convergence, since $1_n(t) \leq 1_{n-1}(t)$ for all $t \geq 0$ and $n \in \mathbb{N}$. □

Remark 3. As illustrated in the following example, we have been unsuccessful in using (15) to obtain a closed form expression for F even for simple choices of γ. Fortunately, the recursion (14) provides a useful numerical approximation to F. As the integral in (13) with $f = 1_{n-1}$ quickly becomes difficult to evaluate analytically as n increases, we compute the integral numerically, using a quadrature rule. □

Example 4. (*Unmarked process*) Consider Example 1 with $\beta = 1$. Then (15) is equivalent to

$$\int_0^t F(s)\mathrm{e}^s\,\mathrm{d}s = \frac{\mathrm{e}^t}{\alpha}\ln(\mathrm{e}^\alpha F(t))$$

which is not analytically solvable. □

4.2. Monotonicity properties and convergence results

As established in Theorem 1 below, many other approximations of F than 1_n exist, and the rate of convergence may be geometric with respect to different norms. First we notice that certain monotonicity properties are fulfilled by φ, where we for functions $f : [0,\infty) \mapsto [0,1]$ set $f_0 = \varphi^{[0]}(f) = f$ and define recursively $f_n = \varphi^{[n]}(f) = \varphi(f_{n-1})$, $n \in \mathbb{N}$. Note that $F_n = F$ for all $n \in \mathbb{N}_0$. As $1_n = \varphi^{[n]}(1)$ is decreasing towards the c.d.f. F, cases where G is a c.d.f. and G_n increases to F are of particular interest.

Lemma 2. *For any $f, g \in \mathcal{C}$, we have that*

$$f \le g \;\Rightarrow\; f_n \le g_n, \qquad n \in \mathbb{N}, \tag{16}$$

$$f \le \varphi(f) \;\Rightarrow\; f_n \text{ is non-decreasing in } n, \tag{17}$$

$$f \ge \varphi(f) \;\Rightarrow\; f_n \text{ is non-increasing in } n. \tag{18}$$

Proof. We obtain immediately (16) from (13) when $n = 1$, whereby (16) follows by induction. Thereby (17) and (18) follow. □

Theorem 1. *With respect to the supremum norm $\|f\|_\infty = \sup_{t\ge 0} |f(t)|$, φ is a contraction on $\mathcal{C}$, that is, for all $f, g \in \mathcal{C}$ and $n \in \mathbb{N}$, we have that $f_n, g_n \in \mathcal{C}$ and*

$$\|\varphi(f) - \varphi(g)\|_\infty \le \bar{\nu}\|f - g\|_\infty. \tag{19}$$

Further, F is the unique fixpoint,

$$\|F - f_n\|_\infty \to 0 \qquad \text{as } n \to \infty, \tag{20}$$

and

$$\|F - f_n\|_\infty \le \frac{\bar{\nu}^n}{1-\bar{\nu}}\|\varphi(f) - f\|_\infty, \tag{21}$$

where $\|\varphi(f) - f\|_\infty \le 1$. Furthermore, if $f \le \varphi(f)$ (or $f \ge \varphi(f)$), then f_n converges to F from below (above).

Proof. Let $f, g \in \mathcal{C}$. Recall that by the mean value theorem (e.g. Theorem 5.11 in [1]), for any real numbers x and y, $\mathrm{e}^x - \mathrm{e}^y = (x-y)\mathrm{e}^{z(x,y)}$, where $z(x,y)$ is a real number between x and y. Thus by (13),

$$\|\varphi(f) - \varphi(g)\|_\infty = \sup_{t\ge 0}\left|\mathrm{E}\left[\mathrm{e}^{-\nu}\mathrm{e}^{c(t,f,g)}\int_0^t (f(t-s) - g(t-s))\gamma(s,Z)\mathrm{d}s\right]\right| \tag{22}$$

where $c(t,f,g)$ is random variable between $\int_0^t f(t-s)\gamma(s,Z)\,\mathrm{d}s$ and $\int_0^t g(t-s)\gamma(s,Z)\,\mathrm{d}s$. Since $f,g\le 1$, we obtain $\mathrm{e}^{c(t,f,g)}\le \mathrm{e}^{\nu}$, cf. (2). Consequently,

$$\begin{aligned}\|\varphi(f)-\varphi(g)\|_\infty &\le \sup_{t\ge 0}\left|\mathrm{E}\left[\int_0^t (f(t-s)-g(t-s))\gamma(s,Z)\,\mathrm{d}s\right]\right| \\ &\le \mathrm{E}\left[\int_0^\infty \|f-g\|_\infty\,\gamma(s,Z)\,\mathrm{d}s\right] = \bar{\nu}\,\|f-g\|_\infty.\end{aligned}$$

Thereby (19) is verified. Since $\mathcal{C}$ is complete (see e.g. Theorem 3.11 in [25]), it follows from the fixpoint theorem for contractions (see e.g. Theorem 4.48 in [1]) that the contraction has a unique fixpoint. By (15), this is F.

Since $f\in\mathcal{C}$ implies $\varphi(f)\in\mathcal{C}$, we get by induction that $f_n\in\mathcal{C}$. Hence, using (15), (19) and induction,

$$\|f_n-F\|_\infty = \|\varphi(f_{n-1})-\varphi(F)\|_\infty \le \bar{\nu}\|f_{n-1}-F\|_\infty \le \bar{\nu}^n\|f-F\|_\infty, \qquad n\in\mathbb{N}. \tag{23}$$

Since $\bar{\nu}<1$, (20) is obtained.

For similar reasons as in (23),

$$\|f_n-f_{n-1}\|_\infty \le \bar{\nu}^{n-1}\|f_1-f\|_\infty, \qquad n\in\mathbb{N}. \tag{24}$$

Further, by (20),

$$\|F-f\|_\infty = \lim_{m\to\infty}\|f_m-f\|_\infty.$$

So by the triangle inequality and (24),

$$\begin{aligned}\|F-f\|_\infty &\le \lim_{m\to\infty}\big(\|f_1-f\|_\infty+\|f_2-f_1\|_\infty+\cdots+\|f_m-f_{m-1}\|_\infty\big) \\ &\le \lim_{m\to\infty}\|f_1-f\|_\infty\big(1+\bar{\nu}+\cdots+\bar{\nu}^{m-1}\big) = \|f_1-f\|_\infty/\big(1-\bar{\nu}\big),\end{aligned}$$

cf. (2). Combining this with (23), we obtain (21). Finally, if $f\le\varphi(f)$ (or $f\ge\varphi(f)$) then by (17) (or (18)) and (20), f_n converges from below (or above). □

Similar results to those in Theorem 1 but for the L^1-norm are established in [17]. The following remark and proposition show how to find upper and lower bounds of F in many cases.

Remark 4. Consider a function $f\in\mathcal{C}$. The condition $f\le\varphi(f)$ or $f\ge\varphi(f)$ is satisfied for the extreme cases $f=0$ or $f=1$. The upper bound $f=1$ is useful in the following sections, but the lower bound $f=0$ is a too small function for our

purposes; if we require that $\mathrm{E}L < \infty$, cf. Remarks 1 (in fact we use only $f = 0$ when producing the right plot in Figure 4). To obtain a more useful lower bound, observe that $f \le \varphi(f)$ implies $f \le F < 1$, cf. (5) and Theorem 1. If $f < 1$, then a sufficient condition for $f \le \varphi(f)$ is

$$\frac{1}{\bar{\nu}} \ge \frac{\int_0^t (1 - f(t-s))\bar{h}(s)\,\mathrm{d}s + \int_t^\infty \bar{h}(s)\,\mathrm{d}s}{1 - f(t)}, \qquad t \ge 0. \tag{25}$$

This follows readily from (7) and (13), using that $e^x \ge 1 + x$.

The case where f in (25) is closest to F happens when f is a c.d.f. G and we have equality in (25). Equivalently, G satisfies the renewal equation

$$G(t) = 1 - \bar{\nu} + \bar{\nu} \int_0^t G(t-s)\bar{h}(s)\,\mathrm{d}s, \qquad t \ge 0,$$

which has the unique solution

$$G(t) = 1 - \bar{\nu} + \sum_{n=1}^{\infty} (1-\bar{\nu})\bar{\nu}^n \int_0^t \bar{h}^{*n}(s)\,\mathrm{d}s, \qquad t \ge 0, \tag{26}$$

where $*n$ denotes convolution n times, cf. Theorem IV2.4 in [2]. In other words, G is the c.d.f. of $\bar{R}_1 + \cdots + \bar{R}_K$ (setting $\bar{R}_1 + \cdots + \bar{R}_K = 0$ if $K = 0$), where $K, \bar{R}_1, \bar{R}_2, \ldots$ are independent random variables, each $\bar{R}_i$ has density $\bar{h}$, and K follows a geometric density $(1-\bar{\nu})\bar{\nu}^n$. Interestingly, this geometric density is equal to $\mathrm{E}S_n/\mathrm{E}S$, cf. (4).

The next proposition shows that in many situations $G \le \varphi(G)$ when G is an exponential c.d.f. with a sufficiently large mean. In such cases F has no heavier tails than such an exponential distribution. □

Denote by

$$\mathcal{L}(\theta) = \int_0^\infty \mathrm{e}^{\theta t}\bar{h}(t)\,\mathrm{d}t, \qquad \theta \in \mathbb{R},$$

the Laplace transform of $\bar{h}$.

Proposition 3. *If $G(t) = 1 - \mathrm{e}^{-\theta t}$ for $t \ge 0$, where $\theta > 0$ and $\mathcal{L}(\theta) \le 1/\bar{\nu}$, then $G \le \varphi(G)$.*

Proof. Inserting $f = G$ into the right side of (25) we obtain

$$\int_0^t \mathrm{e}^{\theta s}\bar{h}(s)\,\mathrm{d}s + \mathrm{e}^{\theta t}\int_t^\infty \bar{h}(s)\,\mathrm{d}s.$$

Since this is an increasing function of $t > 0$, (25) is satisfied if and only if $\mathcal{L}(\theta) \le 1/\bar{\nu}$. □

Note that Proposition 3 always applies for sufficiently small $\theta > 0$ except in the case where $\bar{h}$ is heavy-tailed in the sense that $\mathcal{L}(\theta) = \infty$ for all $\theta > 0$.

4.3. Examples

For Examples 5 and 6 below, we let

$$G(t) = 1 - \mathrm{e}^{-\theta t}, \qquad t \geq 0, \tag{27}$$

be the exponential c.d.f. with parameter $\theta > 0$.

Example 5. (*Unmarked process*) For the case in Example 1, $\mathcal{L}(\theta) = \beta/(\beta-\theta)$ if $\theta < \beta$, and $\mathcal{L}(\theta) = \infty$ otherwise. Interestingly, for "the best choice" $\theta = \mathcal{L}^{-1}(1/\bar{\nu}) = \beta(1-\alpha)$, (27) becomes the c.d.f. for R times $\mathrm{E}S$, which is easily seen to be the same as the c.d.f. in (26).

Figure 4 (at left) shows 1_n and G_n when $\theta = \beta(1-\alpha)$ and $(\alpha, \beta) = (0.9, 1)$. The convergence of 1_n and G_n (with respect to $\|\cdot\|_\infty$) and the approximate form of F are clearly visible. Since G is a c.d.f. and $G_{n+1} \geq G_n$, we have that G_n is a c.d.f. Figure 4 (at middle) shows the density $F'(t)/(1-F(0))$ $(t > 0)$ approximated by $[1_n'(t)/(1-1_n(0)) + G_n'(t)/(1-G_n(0))]/2$ when $n = 50$ (in which case $1_n'(t)/(1-1_n(0))$ and $G_n'(t)/(1-G_n(0))$ are effectively equal). As shown in the plot, the density is close to the exponential density with the same mean, but the tail is slightly thicker. □

Example 6. (*Birth-death process*) For the case in Example 2,

$$\mathcal{L}(\theta) = \mathrm{E}\int_0^Z \mathrm{e}^{\theta s}/\mathrm{E}Z \,\mathrm{d}s = \frac{\mathcal{L}_Z(\theta) - 1}{\theta \mathrm{E}Z}$$

where $\mathcal{L}_Z(\theta) = \mathrm{E}\mathrm{e}^{\theta Z}$ is the Laplace transform for Z. In the special case where Z is exponentially distributed with mean $1/\beta$, $\mathcal{L}(\theta) = \mathcal{L}_Z(\theta) = \beta/(\beta-\theta)$ is of the same form as in Example 5. Plots of 1_n, G_n, and $[1_n'/(1-1_n(0)) + G_n'/(1-G_n(0))]/2$ for $n = 0, 5, \ldots, 50$ and $(\alpha, \beta) = (0.9, 1)$ are similar to those in Figure 4 (at right and middle) and are therefore omitted. □

Example 7. (*A heavy-tailed distribution for L*) For the case in Example 3, Proposition 3 does not apply as $\mathcal{L}(\theta) = \infty$ for all $\theta > 0$. The c.d.f. in (26) is not known on closed form, since the convolutions are not tractable (in fact this is the case when $\bar{h}$ specifies any known heavy-tailed distribution, including the Pareto, Weibull, lognormal

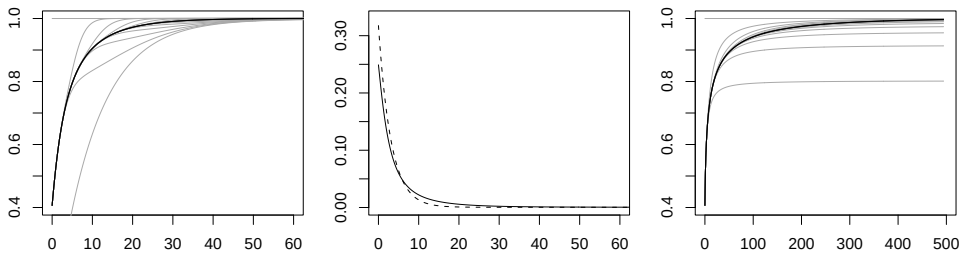

FIGURE 4: Left: 1_n and G_n for $n = 0, 5, \ldots, 50$ in the unmarked case with $\alpha = 0.9$ and $\beta = 1$ (see Example 5); 1_{50} and G_{50} are shown in black to illustrate the approximate form of F whereas the other curves are gray. Middle: The density $[1_n'/(1 - 1_n(0)) + G_n'/(1 - G_n(0))]/2$ when $n = 50$ (solid line) and the exponential density with the same mean (dashed line) Right: as left but for Example 7 with $\alpha = 0.9$ and $\beta = 1$ using 1_n and 0_n as approximations of F.

or loggamma distribution). Nonetheless, it is still possible to get an idea of what F looks like: Figure 4 (at right) shows 1_n and 0_n for $n = 0, 5, \ldots, 50$ in the case $(\alpha, \beta) = (0.9, 1)$. As in Examples 5 and 6, the convergence of 1_n and G_n (where now $G = 0$) and the approximate form of F are clearly visible. However, as indicated in Figure 4 (at right) and verified in [17], $\lim_{t \to 0} G_n(t) < 1$ when $G = 0$, so G_n is not a c.d.f. □

5. Simulation of I_2

To complete the perfect simulation algorithm (Algorithm 2 in Section 3), we need a useful way of simulating I_2. Our procedure is based on a dominating process and the use of coupled upper and lower processes in a similar spirit as in the dominated coupling from the past algorithm in Kendall and Møller [15].

Suppose that $f \in \mathcal{C}$ is on a closed form, $f \leq \varphi(f)$, and (9) is satisfied when we replace F by f (situations where these requirements are fulfilled are considered in Sections 3.2, 4.2 and 4.3). Particularly, if μ is constant and f is a c.d.f., (9) means that f has a finite mean. Now, for $n \in \mathbb{N}_0$, let U_n and L_n denote Poisson processes on $(-\infty, 0)$ with intensity functions $\lambda_n^u(t) = (1 - f_n(-t))\mu(t)$ and $\lambda_n^l(t) = (1 - 1_n(-t))\mu(t)$, respectively. By Theorem 1, λ_n^u is non-increasing and λ_n^l is non-

decreasing in n, and they both converge to λ_2 (geometrically fast with respect to the supremum norm). Consequently, we can use independent thinning to obtain the following sandwiching/funneling property, cf. [15]:

$$\emptyset = L_0 \subseteq L_1 \subseteq L_2 \subseteq \cdots \subseteq I_2 \subseteq \cdots \subseteq U_2 \subseteq U_1 \subseteq U_0. \tag{28}$$

The details are given by the following algorithm.

Algorithm 3. Simulation of I_2:

(i) Generate a realisation $\{(t_1, Z_1), \ldots, (t_k, Z_k)\}$ of U_0, where $t_1 < \ldots < t_k$.

(ii) If $U_0 = \emptyset$, then return $I_2 = \emptyset$ and stop, else generate independent uniform numbers $W_1, \ldots, W_k$ on $[0, 1]$ (independently of U_0), and set $n = 1$.

(iii) For $j = 1, \ldots, k$, assign (t_j, Z_j) to L_n respective U_n if $W_j \lambda_0^u(t_j) \leq \lambda_n^l(t_j)$ respective $W_j \lambda_0^u(t_j) \leq \lambda_n^u(t_j)$.

(iv) If $U_n = L_n$, then return $I_2 = L_n$ and stop, else increase n by 1 and repeat steps (iii)–(iv).

Proposition 4. *Algorithm 3 works correctly and terminates almost surely within finite time.*

Proof. To see this, imagine no matter if $U_0 = \emptyset$ in step (ii) or $U_n = L_n$ in step (iv), we continue to generate (U_1, L_1), (U_2, L_2), etc. Further, add an extra step: For $j = 1, \ldots, k$, assign (t_j, Z_j) to I_2 if and only if $W_j \lambda_n^u(t_j) \leq \lambda_2(t_j)$. Then clearly, because of the convergence properties of λ_n^u and λ_n^l (see the discussion above), (28) is satisfied and conditional on $t_1, \ldots, t_k$,

$$\begin{aligned} \mathrm{P}(L_n \neq U_n \ \forall n \in \mathbb{N}_0) &\leq \sum_{j=1}^{k} \lim_{n \to \infty} \mathrm{P}(W_j \lambda_0^u(t_j) \leq \lambda_n^u(t_j),\ W_j \lambda_0^u(t_j) > \lambda_n^l(t_j)) \\ &= \sum_{j=1}^{k} \mathrm{P}(\lambda_2(t_j) < W_j \lambda_0^u(t_j) \leq \lambda_2(t_j)) = 0. \end{aligned}$$

Thus almost surely Algorithm 3 terminates within finite time and the output equals I_2. □

Remark 5. We compute 1_n and f_n numerically, using a quadrature rule, cf. Remark 3. After step (i) in Algorithm 3, we let the last quadrature point be given by $-t_1$ (since

we do not need to calculate $1_n(t)$ and $f_n(t)$ for $t > -t_1$). Since we have to calculate 1_n and f_n recursively for all $n = 0, 1, 2, \ldots$ until Algorithm 3 terminates, there is no advantage in using a doubling scheme for n like in the Propp-Wilson algorithm [23]. □

Example 8. (*Birth-death process)* We have checked our computer code for Algorithms 2 and 3 by comparing with results produced by another perfect simulation algorithm: Consider the case in Example 2 when $\mu(t) = \mu$ is constant and Z is exponentially distributed with mean $1/\beta$. If N denotes the number of events alive at time 0, we have the following detailed balance condition for its equilibrium density π_n:

$$\pi_n(\mu + \alpha\beta n) = \pi_{n+1}\beta(n+1), \qquad n \in \mathbb{N}_0.$$

This density is well-defined, since $\lim_{n\to\infty} \pi_{n+1}/\pi_n = \alpha < 1$. Now, choose $m \in \mathbb{N}_0$ and $\epsilon \geq 0$ such that $a = \alpha + \epsilon < 1$ and $\pi_{n+1}/\pi_n \leq a$ whenever $n \geq m$. If $\mu \leq \alpha\beta$, we can take $\epsilon = m = 0$; otherwise we can use $m \geq (\mu - \alpha\beta)/(\beta\epsilon)$ for some $\epsilon > 0$. Define an unnormalised density $\pi'_n,\ n \in \mathbb{N}_0$, by $\pi'_n = \pi_n/\pi_0$ if $n \leq m$, and $\pi'_n = a^{n-m}\pi_m/\pi_0$ otherwise. We can easily sample from π'_n by inversion, cf. [24], since we can calculate

$$\sum_0^\infty \pi'_n = \sum_0^m \pi_n/\pi_0 + \frac{a}{1-a}\,\pi_m/\pi_0.$$

Then, since $\pi'_n \geq \pi_n/\pi_0$, we can sample N from π_n by rejection sampling, cf. [24]. Furthermore, conditional on $N = n$, we generate n independent marks $Z'_1, \ldots, Z'_n$ which are exponentially distributed with mean $1/\beta$ (here we exploit the memoryless property of the exponential distribution). Finally, we simulate the marked Hawkes process with events in $(0, t_+]$, using the conditional intensity

$$\lambda'(t) = \mu + \alpha\beta\left(\sum_{i=1}^n \mathbf{1}[t < Z'_i] + \sum_{0<t_i<t} \mathbf{1}[t < t_i + Z_i]\right).$$

We have implemented this algorithm for comparison with our algorithm. Not surprisingly this algorithm is a lot faster than our perfect simulation algorithm (roughly 1200 times as fast in the case $\alpha = 0.9$, $\beta = \mu = 1$, and $t_+ = 10$), since it exploits the fact that we know the stationary distribution in this special case. □

6. Extensions and open problems

Except for the heavy-tailed case, our perfect simulation algorithm is feasible in the examples we have considered. In heavy-tailed cases, we can only say something about the approximate form of F, cf. Example 7.

Many of our results and algorithms can be modified if we slightly extend the definition in Section 1 of a marked Hawkes process as follows: For any event t_i with associated mark Z_i, let n_i denote the number of (first generation) offspring generated by (t_i, Z_i), and suppose that n_i conditional on Z_i is not necessarily Poisson distributed, but n_i is still conditionally independent of t_i and the previous history. A particular simple case occurs when n_i is either 1 or 0, where $\bar{p} = \mathrm{EP}(n_i = 1|Z_i)$ is assumed to be strictly between 0 and 1 (here $\bar{p}$ plays a similar role as $\bar{\nu}$ introduced in Section 4). Then we redefine φ by

$$\varphi(f)(t) = 1 - \bar{p} + \bar{p} \int_0^t f(t-s)\bar{h}(s)\,\mathrm{d}s$$

where now

$$\bar{h}(s) = \mathrm{E}(p(Z)h(s, Z))/\bar{p}.$$

Since φ now is linear, the situation is much simpler. For example, F is given by G in (26) (with $\bar{\nu}$ replaced by $\bar{p}$).

For applications such as in seismology (see e.g. [22]), extensions of our results and algorithms to the case of predictable marks are both important and challenging. However, F becomes much more complicated, and it seems e.g. hard to extend the proof in Proposition 1 because the self-similarity property (Section 2.2) is lost.

Extensions to non-linear Hawkes processes [3, 7] would also be interesting. Again things become complicated, since a non-linear Hawkes process is not even a Poisson cluster process.

Another extension of practical relevance is to consider a non-Poisson immigrant process, e.g. a Markov or Cox process. The results in Section 4 do not depend on the choice of immigrant process, and the straightforward simulation algorithm (Algorithm 1 in Section 3.1) applies provided it is feasible to simulate the immigrants on $[t_-, t_+)$. However, the perfect simulation algorithm relies much on the assumption that the immigrant process is Poisson.

Finally, we notice that it would be interesting to extend our ideas to spatial Hawkes processes, cf. [18] and [19].

Acknowledgements

We are grateful to Søren Asmussen, Svend Berntsen, Horia Cornean, Martin Jacobsen, Arne Jensen, and Giovanni Luca Torrisi for helpful discussions. The research of Jesper Møller was supported by the Danish Natural Science Research Council and the Network in Mathematical Physics and Stochastics (MaPhySto), funded by grants from the Danish National Research Foundation.

References

[1] Apostol, T. M. (1974). *Mathematical Analysis.* Addison-Wesley, Reading.

[2] Asmussen, S. (1987). *Applied Probability and Queues.* Wiley, Chichester.

[3] Brémaud, P. and Massoulié, L. (1996). Stability of nonlinear Hawkes processes. *Ann. Prob.* **24,** 1563–1588.

[4] Brémaud, P., Nappo, G. and Torrisi, G. (2002). Rate of convergence to equilibrium of marked Hawkes processes. *J. Appl. Prob.* **39,** 123–136.

[5] Brix, A. and Kendall, W. (2002). Simulation of cluster point processes without edge effects. *Adv. Appl. Prob.* **34,** 267–280.

[6] Chornoboy, E. S., Schramm, L. P. and Karr, A. F. (2002). Maximum likelihood identification of neural point process systems. *Adv. Appl. Prob.* **34,** 267–280.

[7] Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes, Volume I: Elementary Theory and Methods* 2nd ed. Springer, New York.

[8] Dwass, M. (1969). The total progeny in a branching process and a related random walk. *J. Appl. Prob.* **6,** 682–686.

[9] Hawkes, A. G. (1971). Point spectra of some mutually exciting point processes. *J. Roy. Statist. Soc. Ser. B* **33,** 438–443.

[10] Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58,** 83–90.

[11] Hawkes, A. G. (1972). Spectra of some mutually exciting point processes with associated variables. In *Stochastic Point Processes*. ed. P. A. W. Lewis. Wiley, New York, pp. 261–271.

[12] Hawkes, A. G. and Adamopoulos, L. (1973). Cluster models for earthquakes – regional comparisons. *Bull. Int. Statist. Inst.* **45,** 454–461.

[13] Hawkes, A. G. and Oakes, D. (1974). A cluster representation of a self-exciting process. *J. Appl. Prob.* **11,** 493–503.

[14] Jagers, P. (1975). *Branching Processes with Biological Applications*. John Wiley & Sons, London.

[15] Kendall, W. S. and Møller, J. (2000). Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Adv. Appl. Prob.* **32,** 844–865.

[16] Møller, J. (2003). Shot noise Cox processes. *Adv. Appl. Prob.* **35,** 614–640.

[17] Møller, J. and Rasmussen, J. G. (2004). Approximate simulation of Hawkes processes. In preparation.

[18] Møller, J. and Torrisi, G. L. (2004). Generalised shot noise Cox processes. *Technical report* R-2004-07. Department of Mathematical Sciences, Aalborg University.

[19] Møller, J. and Torrisi, G. L. (2004). Spatial Hawkes processes. In preparation.

[20] Møller, J. and Waagepetersen, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman & Hall, Boca Raton, Florida.

[21] Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *J. Amer. Statist. Assoc.* **83,** 9–27.

[22] Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Ann. Inst. Statist. Math.* **50,** 379–402.

[23] Propp, J. G. and Wilson, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms* **9,** 223–252.

[24] Ripley, B. D. (1987). *Stochastic Simulation.* Wiley, New York.

[25] Rudin, W. (1987). *Real and Complex Analysis.* McGraw-Hill, New York.

[26] Vere-Jones, D. and Ozaki, T. (1982). Some examples of statistical inference applied to earthquake data. *Ann. Inst. Statist. Math.* **34,** 189–207.